Reply to Guy et al.: Support for a bottleneck in the 2011 Escherichia coli O104:H4 outbreak in Germany

Grad, Yonatan H.; Lipsitch, Marc; Griggs, Allison D.; Haas, Brian J.; Shea, Terrance P.; McCowan, Caryn; Montmayeur, Anna; FitzGerald, Michael; Wortman, Jennifer R.; Krogfelt, Karen Angeliki

Total number of authors:
19

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1209419110

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

LETTER

Reply to Guy et al.: Support for a bottleneck in the 2011 *Escherichia coli* O104:H4 outbreak in Germany

In our paper (1), we analyzed isolates from the *Escherichia coli* O104:H4 outbreaks in Germany and France in May to July 2011. We concluded that, although the German outbreak was larger, the German isolates represent a clade within the greater diversity of the French outbreak. We proposed several hypotheses to explain these findings, including that the lineage leading to the German outbreak went through a narrow bottleneck that purged diversity.

Guy et al. (2) report the genomes of eight additional *E. coli* O104:H4 isolates sampled from the German outbreak. By focusing on the numbers of SNPs in their samples, they suggest that the German outbreak is more diverse than we reported and is similar to the French outbreak.

In fact, Guy et al.'s data (2) strongly support our conclusion that the German outbreak represents a clade within the diversity described by the French outbreak. We analyzed the raw data [kindly supplied by Guy et al. (2)] using the same SNP-calling approach described in our previous work to allow for an accurate comparison unbiased by differences in methods (1); the analysis yields the same tree structure as that described by Guy et al. (2), with a slightly different set of SNPs and branch lengths (Fig. 1*A* and Table 1).

The tree shows that all the German outbreak isolates belong to a single clade with a star phylogeny, with one exception (E92/11). The star phylogeny is consistent with a single point source and population expansion. By contrast, the French outbreak isolates have branching structure, indicative of a distinct pattern of diversity.

Our conclusion is further supported by subsequent data we have obtained in collaboration with the Robert Koch Institute and Pasteur Institute, including (*i*) sequencing of an additional 10 outbreak isolates from Germany and seven from France (Fig. 1*B*) and (*ii*) genotyping of 47 more isolates from the German outbreak, all of which have the SNPs that define the German outbreak clade in our original analysis (sites 1568661 and 2252380) and none of which have SNPs we identified in the French outbreak.

The sole exception among the 22 fully sequenced (Fig. 1*B*) and 47 genotyped German outbreak isolates analyzed here is E92/11, which clusters with isolates from the French outbreak. This anomalous isolate may reflect an incomplete bottleneck in the German outbreak, such that contaminating bacteria survived the bottleneck at different frequencies. Alternatively, the sample [which comes from an infected individual who traveled May 7–10, 2011, in Germany, according to data provided by Guy et al. (2)] may reflect exposure to home-grown sprouts, rather than sprouts from the farm implicated as the major source of the outbreak (3), or exposure relatively early in the outbreak (4), predating the bottleneck. Discriminating among these hypotheses requires additional epidemiological data.

We agree with Guy et al. (2) that greater sampling can enhance insight into outbreak dynamics, but note that interpretation of the resulting data requires integration of phylogenetic and epidemiological relationships. In this case, the additional data support the hypothesis of a bottleneck in the German *E. coli* O104:H4 outbreak.

***Yonatan H. Grad*[a,b], *Marc Lipsitch*[b,c], *Allison D. Griggs*[d], *Brian J. Haas*[d], *Terrance P. Shea*[d], *Caryn McCowan*[d], *Anna Montmayeur*[d], *Michael FitzGerald*[d], *Jennifer R. Wortman*[d], *Karen A. Krogfelt*[e], *Edouard Bingen*[f,g], *François-Xavier Weill*[h], *Erhard Tietze*[i], *Antje Flieger*[i], *Eric S. Lander*[d,j,k,1], *Chad Nusbaum*[d], *Bruce W. Birren*[d], *Deborah T. Hung*[a,d,l,m], *and William P. Hanage*[b]**

[a]*Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115;* [b]*Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115;* [c]*Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115;* [d]*Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142;* [e]*Department of Microbial Surveillance and Research, Statens Serum Institute, 2300 Copenhagen, Denmark;* [f]*Laboratoire Associé au Centre National de Référence des Escherichia coli et Shigella, Service de Microbiologie, Hôpital Robert Debré, Assistance Publique–Hôpitaux de Paris, 75019 Paris, France;* [g]*Université Paris–Diderot, Sorbonne Paris Cité, 75505 Paris, France;* [h]*Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Centre National de Référence des Escherichia coli et Shigella, 75015 Paris, France;* [i]*Division of Bacterial Infections and National Reference Centre for Salmonella and Other Enteric Bacterial Pathogens, Robert Koch-Institut, Burgstr. 37, D-38855 Wernigerode, Germany;* [j]*Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139;* [k]*Department of Systems Biology, Harvard Medical School, Boston, MA 02115;* [l]*Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; and* [m]*Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114*

1. Grad YH, et al. (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci USA* 109(8):3065–3070.
2. Guy L, et al. (2012) Genomic diversity of the 2011 European outbreaks of *Escherichia coli* O104:H4. *Proc Natl Acad Sci USA* 109:E3627–E3628.
3. Buchholz U, et al. (2011) German outbreak of Escherichia coli O104:H4 associated with sprouts. *N Engl J Med* 365(19):1763–1770.
4. Frank C, et al.; HUS Investigation Team (2011) Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. *N Engl J Med* 365(19):1771–1780.

Author contributions: Y.H.G., M.L., E.B., F.-X.W., E.T., A.F., D.T.H., and W.P.H. designed research; Y.H.G., A.D.G., B.J.H., T.P.S., C.M., A.M., M.F., J.R.W., C.N., and B.W.B. performed research; K.A.K., E.B., F.-X.W., E.T., and A.F. contributed new reagents/analytic tools; Y.H.G., M.L., A.D.G., B.J.H., T.P.S., C.M., A.M., M.F., J.R.W., F.-X.W., E.S.L., C.N., B.W.B., D.T.H., and W.P.H. analyzed data; and Y.H.G., M.L., F.-X.W., E.S.L., D.T.H., and W.P.H. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: lander@broadinstitute.org.

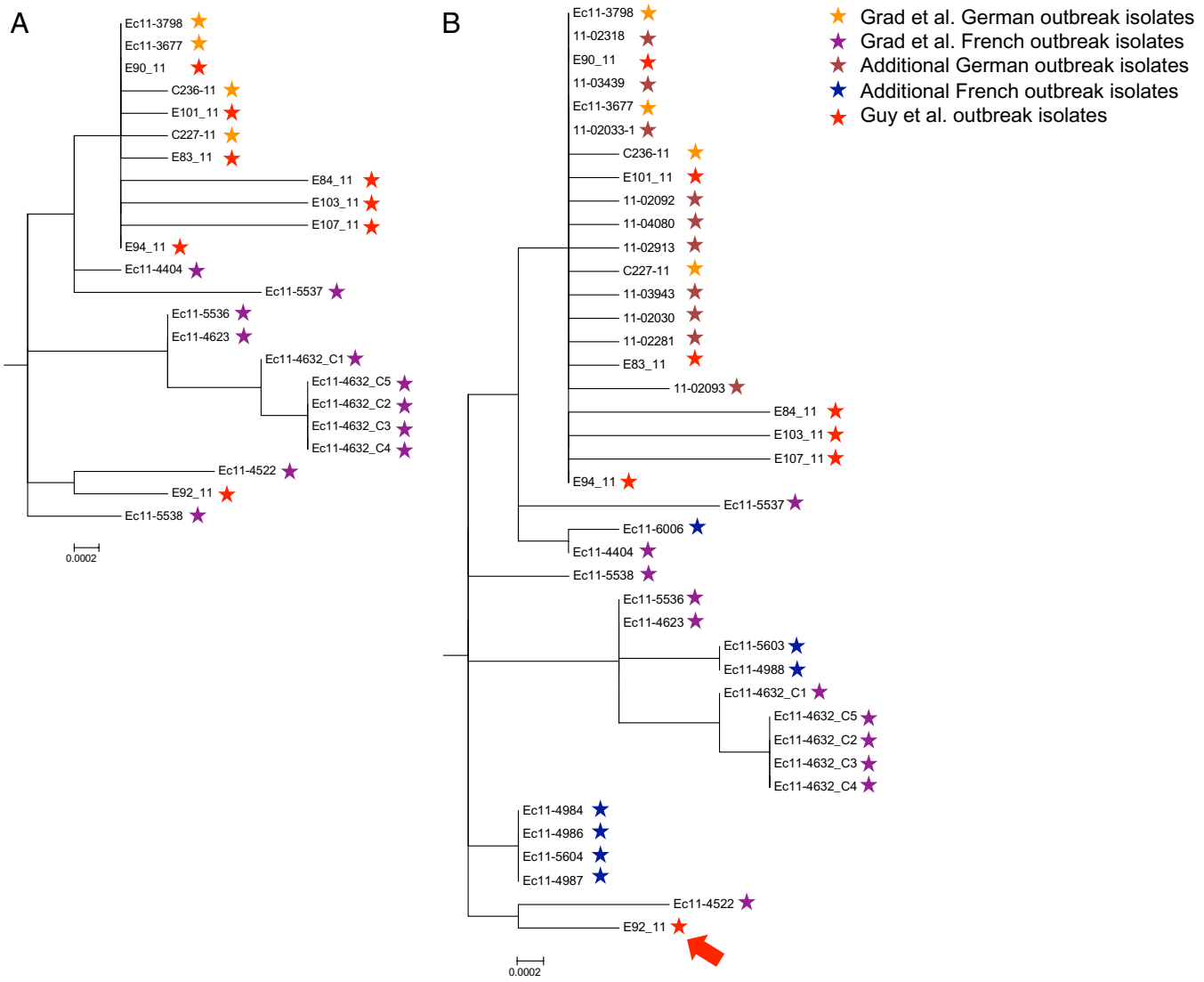

Fig. 1. Maximum likelihood trees of isolates derived from French and German outbreaks, including isolates from Grad et al. (2) and the letter from Guy et al. (1) using SNPs as predicted by our previously described algorithm (2) (*A*) and additional isolates from the German outbreak provided by the Robert Koch Institute (Germany) and from the French outbreak provided by the Pasteur Institute (France) (*B*). The arrow indicates the isolate E92-11, which is the only German outbreak isolate to cluster outside of the clade defined by the remainder of the German outbreak isolates. Further details are provided in the text.

Table 1. SNPs identified by applying the algorithm of Grad et al. (2) to sequence data of Guy et al. (1)

Isolate	Reported SNPs by Guy et al. (1)	SNPs predicted by methods of Grad et al. (2)
E92/11	1262666, 1568661, 2252380, 2564789, 3089339	1262666, 1568661, 2252380, 2564789, 3089339
E83/11	218926	218926
E107/11	4612347	4612347, 1227036*, 2803156*, 2803157*
E103/11	1048209*, 1368969, 1583232, 3170812, 3617762	1368969, 1583232, 3170812, 3617762
E84/11	750858, 4073851, 4613711, 5143640	750858, 4073851, 4613711, 5143640
E101/11	4934415	4934415
E90/11	4253096*	—

Applying the same algorithm as in our paper (2) to Guy et al.'s sequence data (1) identifies 16 of the 18 SNPs called by Guy et al. and predicts three additional SNPs (differences noted by asterisks). Dash indicates no SNPs were predicted in this isolate. Differences in SNP predictions are attributable to different filtering criteria in the SNP prediction pipelines. In addition, Guy et al. (1) express concern that our paper may not have been justified in regarding 54 potential SNPs in TY2482 as likely sequencing errors. At the time of the preparation of our paper, the raw data for TY2482 were unavailable, making SNP prediction in this genome impossible. The raw data have since become available. Analyses of the data by Guy et al. (1) and by us predict that only two of the 54 potential SNPs are valid.